

**Enrichissement de la base de données COLTRAIT afin  
de lier les traits morphologiques des collemboles  
(microarthropodes) aux caractéristiques du milieu**

Florian Delabrosse, Sandrine Salmon, Sophie Gachet

► **To cite this version:**

Florian Delabrosse, Sandrine Salmon, Sophie Gachet. Enrichissement de la base de données COLTRAIT afin de lier les traits morphologiques des collemboles (microarthropodes) aux caractéristiques du milieu. 2011. <hal-00607837>

**HAL Id: hal-00607837**

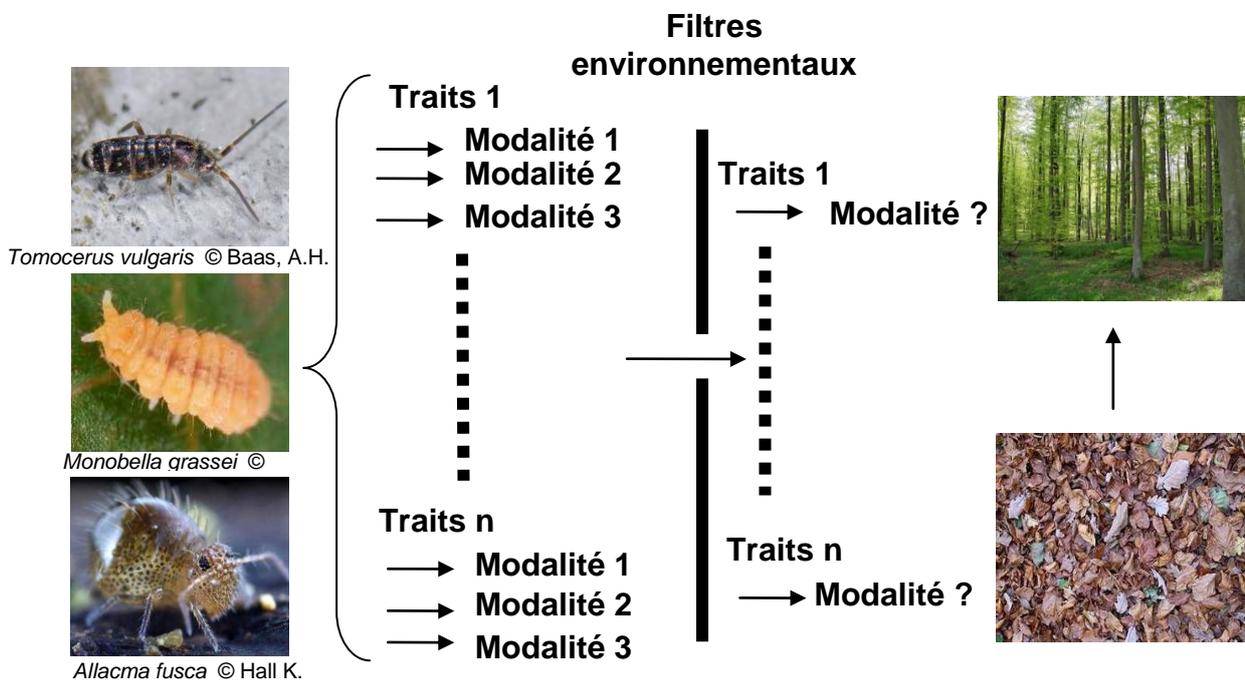
**<https://hal.archives-ouvertes.fr/hal-00607837>**

Submitted on 11 Jul 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Enrichissement de la base de données COLTRAIT afin de lier les traits morphologiques des collemboles (microarthropodes) aux caractéristiques du milieu



**Florian DELABROSSE, Sandrine Salmon (MNHN), Sophie Gachet (IMEP, Marseille)**

Juin 2011

# Sommaire

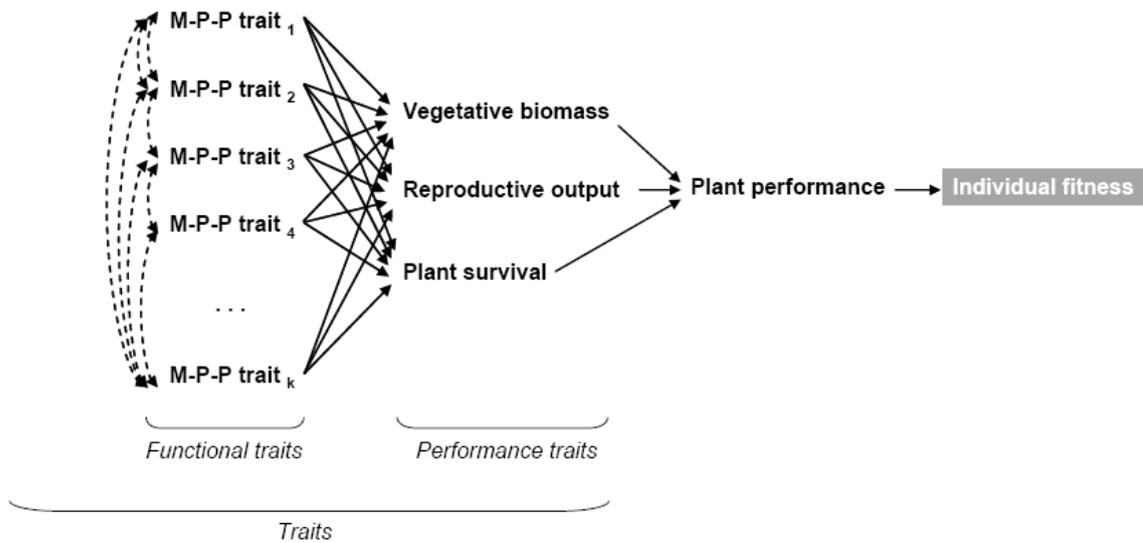
I.	Introduction .....	3
II.	Matériel et méthode.....	6
A.	Sujet de l'étude : les collemboles .....	6
B.	Description de la base de données COLTRAIT .....	6
1.	Structure de la base .....	6
2.	Logiciel Microsoft ACCESS®.....	8
C.	Outils statistiques : .....	8
III.	Résultats .....	9
A.	Apport et modifications pour les différentes tables de COLTRAIT .....	9
1.	Tables « descriptif-relevés » .....	9
2.	Les données d'observation de collemboles : .....	12
3.	Les données d'observation de la végétation.....	13
4.	Table « Traits » .....	14
B.	Distributions et relations entre les traits morphologiques et physiologiques.....	16
1.	Relations entre les traits quantitatifs .....	16
2.	Relations entre les variables traits qualitatives .....	17
C.	Relation entre les traits écologiques et morphologiques : effet de la fermeture du milieu.....	18
1.	Analyses sur les 39 espèces les mieux représentées dans la base : .....	20
2.	Jeu de données avec 34 espèces choisies .....	21
IV.	Discussion .....	22
V.	Conclusion.....	27
VI.	Bibliographie :.....	28
	Annexe 1 : Description des tables et champs de COLTRAIT .....	31
	Annexe 2 : Jeu de données 39 espèces (>15 sources) .....	33
	Annexe 3 : Jeu de données 34 espèces choisies .....	34
	Annexe 4 : Schéma de la structure du tableau de traits aidant au renseignement de la table ..	35
	Annexe 5 : Relation entre les traits quantitatifs .....	36
	Annexe 6 : Analyses sur les 39 espèces les mieux représentées dans la base.....	37
	Annexe 7 : Analyses faites sur le jeu de données de 34 espèces choisies.....	38
	Résumé et Abstract.....	40

## I. Introduction

Les collemboles (arthropodes) constituent l'un des taxons les plus fréquents et abondants de la faune du sol. Ce taxon présente une forte diversité taxonomique et morphologique. La taxonomie des collemboles subit d'ailleurs encore de forts remaniements, notamment au niveau des genres. Ce taxon a également un rôle écologique majeur dans le cycle de la matière et il est aussi utilisé comme indicateur de pollution (survie et reproduction de *Folsomia candida*, norme ISO...)

Les communautés de collemboles observées dans leur milieu forment un « pattern » (patron d'organisation). Un pattern est la représentation d'un assemblage d'espèces données dans un milieu donné. Ce pattern peut être considéré comme résultant de processus aléatoires ou au contraire et c'est le sujet de notre étude, être le fruit de règles, dites règles d'assemblage. Cette notion formalisée par Weither *et al.* en 1995 se base sur les mécanismes de relation espèces-milieu et espèces-espèces. Ainsi l'assemblage d'espèces observé dans une communauté est le fruit : 1) de l'adéquation des besoins des espèces (nourriture, habitat, tolérance au milieu etc.) avec le milieu qu'elle occupe 2) de la satisfaction de ces besoins via la compétition pour les ressources avec les autres espèces. Il existe deux approches pour étudier ces assemblages : l'approche du pattern spécifique d'une communauté, et l'approche du pattern de traits d'une communauté. La seconde relie les traits observés chez les espèces aux deux points précédemment cités : les traits d'adaptation au milieu et les traits relatifs à la « fitness » de l'individu (capacité d'un individu à survivre et se reproduire) face aux autres. Cette seconde approche possède plusieurs avantages : elle est généralisable à d'autres sites, ne demande pas d'expertise particulière (si il existe des bases de traits constitués par des experts) dans le taxon étudié, facilite la construction des règles d'assemblage.

Avant d'aller plus loin définissons la notion de traits (d'après Violle *et al.* 2007). Violle décompose le mécanisme qui mène du trait à la fitness de l'individu (ici pour les plantes) ainsi (Fig. 1) :



**Figure 1. Paradigme de la performance chez un individu, d'après Arnold's (1983)**

**M-P-P : morphologique physiologique et phénologique.**

Dans ce raisonnement, le trait mesuré est relié directement ou indirectement à la fitness de l'individu. La recherche de l'existence et des modalités des règles d'assemblages de communautés se fait en premier lieu par l'identification parmi les traits du taxon de ceux qui sont impliqués dans la fitness individuelle. Les traits fonctionnels pertinents sont identifiés par la valeur explicative qu'ils ont de la variabilité observée des communautés étudiées. Autrement dit la question va être de montrer si un caractère M-P-P (morphologique physiologique et phénologique) va être corrélé à la présence ou non des espèces présentant ce caractère dans un milieu donné ; ou plus simplement de déterminer si un ou des traits M-P-P permettent d'expliquer la présence d'une espèce dans un milieu donné.

Pour revenir aux collemboles-mêmes, il a été montré que la distribution de certaines espèces de collemboles était liée à un certain nombre de facteurs environnementaux tels que le pH, les caractéristiques de la matière organique (via le taux de C, N et cations) mais aussi les caractéristiques physiques des humus (Ponge 1993, Salmon et Ponge 1999, Cassagne 2003).

On peut donc se demander si la composition taxonomique des communautés de collemboles dans un milieu donné, peut être expliquée par l'étude des relations entre les caractéristiques environnementales et les traits morphologiques et d'histoire de vie des taxons.

La réponse à cette question nécessite le traitement d'un grand nombre d'observations afin d'avoir des données provenant du plus grand nombre de relevés et de milieux possibles (puissance statistique). Ce traitement a été fait via l'utilisation d'une base de données, nommée COLTRAIT. La création de la base de données COLTRAIT s'est inspirée des

modèles de bases de données chez les végétaux, notamment la base de données BASECO, une base floristique et écologique de la Flore méditerranéenne française (Gachet *et al.* 2005), qui met en liaison des données de relevés de terrain et de traits morphologiques, provenant d'ouvrages de type flores, de données brutes produites par l'équipe qui a créé la base, de clé de détermination et de synopsis.

La création de COLTRAIT répond à plusieurs besoins. En effet, bien qu'il existe déjà de nombreuses bases de données pour d'autres taxons, aucune base n'existe sur les invertébrés du sol. De plus, l'accumulation d'années de recherche, notamment pour le taxon des collemboles, met à disposition un nombre important d'informations et de données, qui n'ont été utilisées que séparément. Il devient donc intéressant d'utiliser le potentiel informatif important que représentent ces données rassemblées pour des études globales et à grande échelle.

La mise en évidence de règles d'assemblage nécessite cette vue d'ensemble afin de produire des résultats généralisables à d'autres sites, pour lesquels nous disposons d'informations sur leur description, de la zone de prospection (l'Europe).

L'objectif de ce stage a été de compléter la base avec des données d'articles scientifiques et de clés d'identification des collemboles, tout en réfléchissant aux outils permettant de faciliter l'extraction des données des références bibliographiques et le renseignement des traits des espèces de collemboles via les caractères taxonomiques aux stades sous-ordre et famille.

Puis dans un second temps, j'ai déterminé des traits écologiques des espèces, notamment les préférendum écologiques de la fermeture/ouverture du milieu pour les espèces, afin de rechercher des relations entre les traits morphologiques et les traits écologiques.

## II. Matériel et méthode

### A. Sujet de l'étude : les collemboles

On différencie les collemboles des autres entognates par les caractéristiques suivantes (site de Frans Janssen : [www.collembola.org](http://www.collembola.org)) :

- Antennes toujours présentes (absentes chez les protoures)
- Cerques absents (organes postérieurs présents chez les diploures)

Ils possèdent également un organe de saut caractéristique, la furca, mais celle-ci peut être absente chez certaines espèces.

Pour aller plus loin dans la morphologie du collembole :

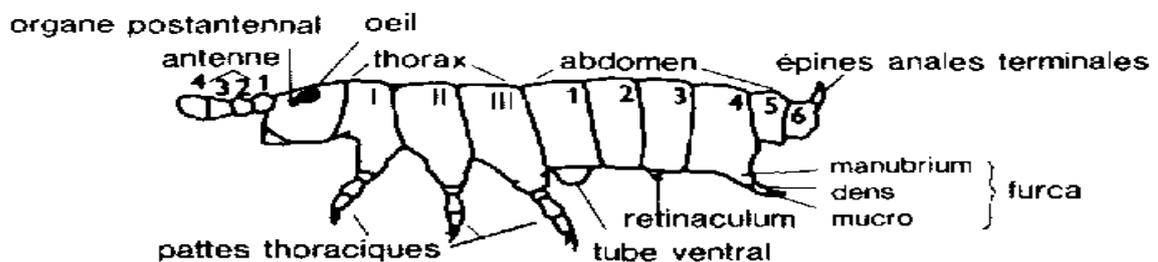


Figure 2. Schéma d'un collembole

Source : <http://www.fao.org>

Il existe quatre sous-ordres de l'ordre « Collembola » et chacun d'entre eux peut être grossièrement associé à des différences morphologiques visibles :

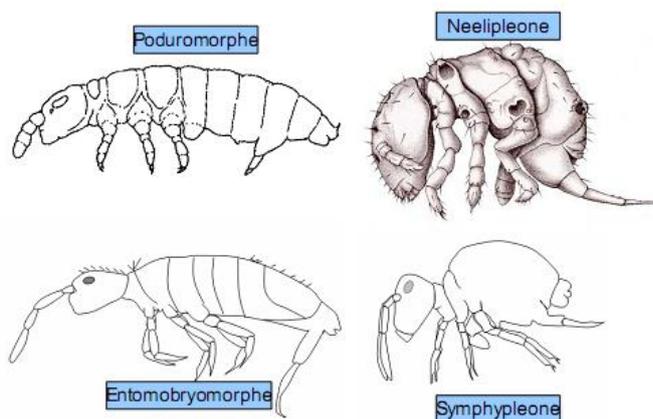


Figure 3. Morphes des quatre sous-ordres de collemboles

### B. Description de la base de données COLTRAIT

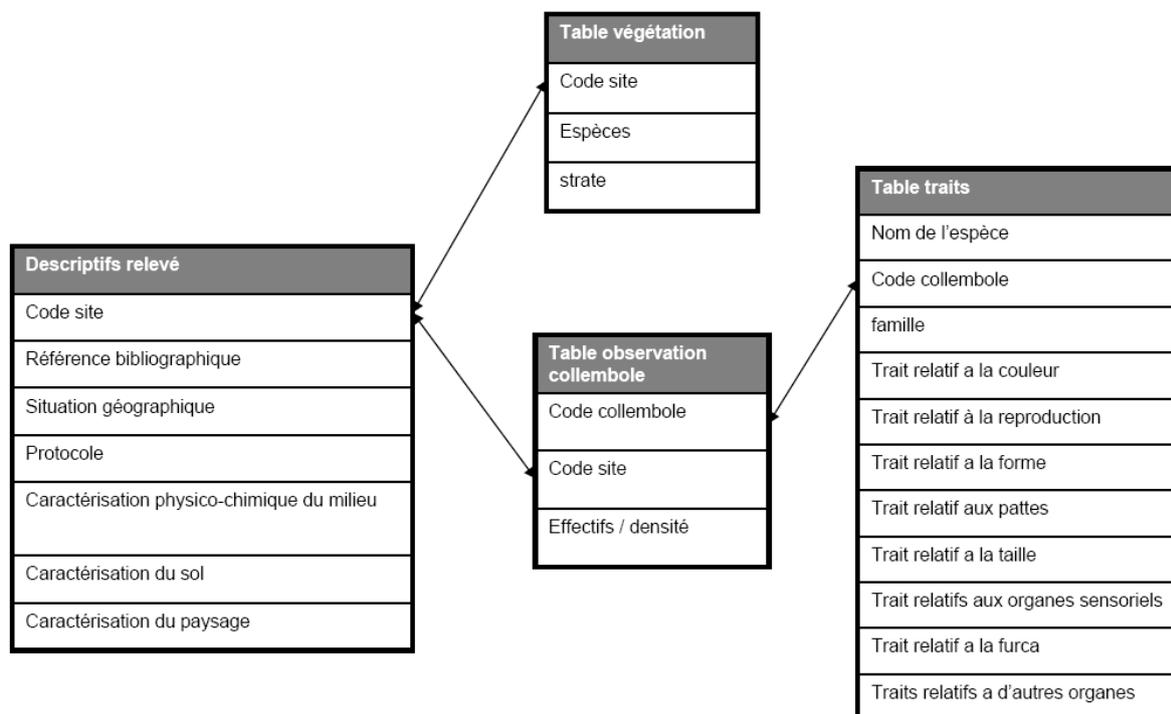
#### 1. Structure de la base

La base COLTRAIT se décompose en quatre tables : une table (113 champs – ou colonnes) relative aux descriptifs de chaque milieu dans lesquels ont été recensés les collemboles, une

table (3 champs) qui décrit la végétation de ces milieux, une table (5 champs) qui fait l'inventaire des espèces de collemboles trouvées dans chaque milieu et une table (44 champs) qui décrit un 25 traits pour chaque espèce de collemboule présente dans au moins un des relevés

Chaque table regroupe un certain type de données (ici par exemple caractéristiques du terrain, ou traits des collemboles). Une table est un tableau. Les colonnes de ce tableau sont appelées « champs ». Chaque champ correspond à une variable (exemple : la longueur du corps ; exception faite des champs : « nom » « codecoll » etc.), cette variable se décline dans les cases de la colonne en modalités (exemple : long, intermédiaire ou court). Les modalités sont les valeurs que peut prendre la variable.

Dans COLTRAIT il existe deux champs « clé primaire » (fig. 3): le « codesite », qui correspond au code du site de chaque relevé et le « codecoll », composé des trois premières lettres du genre suivi des trois premières lettres de l'espèce et qui identifie chaque espèce de façon unique. Ces champs « clé primaire » permettent également de relier les tables entre elles afin de faire correspondre différents type d'informations.



**Figure 4 : Schémas de l'organisation de la base COLTRAIT**

Ainsi, la ligne comprenant par exemple le codecoll « Parnot » (*Parisotoma notabilis*) et décrivant les différents traits de cette espèce dans la table « Traits », est reliée aux lignes

comprenant la mention Parnot dans le champ codecoll de la table contenant les effectifs de collemboles trouvés dans chaque relevé. Ce système relie donc les traits morphologiques de *Parisotoma notabilis* à ses effectifs dans les relevés dans lesquels il a été trouvé. Les informations de ces lignes sont ainsi liées avec un minimum de mémoire utilisée et en conservant une structure particulière pour les deux tables. Les deux tables peuvent être ainsi gérées séparément tout en étant reliées lors de leur modification.

A partir de cette structure, l'on va pouvoir tester les relations entre les deux types d'informations (traits et caractéristiques du milieu).

Les données utilisées notamment pour les relevés de terrain sont limitées à la zone « européenne » délimitée par l'Oural, la Méditerranée l'Atlantique et l'Arctique, qui correspond à l'aire de répartition commune à beaucoup d'espèces européennes.

Comme toute base de données, la base de données COLTRAIT n'est pas totalement aboutie, elle est amenée à évoluer à l'usage et en fonction des documents trouvés.

## 2. Logiciel Microsoft ACCESS®

La base COLTRAIT a été développée sous le logiciel Access (Système de Gestion de Bases de Données Relationnelles).

Ce logiciel permet notamment :

- De gérer un très grand nombre de colonnes et de lignes
- D'importer et d'exporter des données directement depuis Excel
- De formuler des requêtes visant à créer de nouveaux champs interagissant avec d'autres champs préexistant
- De déplacer les champs les uns par rapport aux autres et de les trier tout en conservant la correspondance entre les colonnes tout en maintenant l'unité des lignes.
- De filtrer un certain nombre de lignes selon une modalité d'un ou plusieurs champs

### **C. Outils statistiques :**

Les différents tests et graphiques ont été réalisés à l'aide d'XLSTAT et de R.

Afin de mettre en évidence les associations entre modalités de traits retrouvées le plus fréquemment dans les espèces de la base de données, on a réalisé une Analyse des Correspondances Multiples (ACM) avec les variables qualitatives de la table « Traits ». A

partir de ce résultat on a choisi d'utiliser les variables dont les modalités étaient associées, car on est susceptible de retrouver le même effet du milieu sur celles-ci. La fermeture des milieux où ont été retrouvés les collemboles est la variable environnementale la mieux renseignée dans la table « descriptif-relevés ». De plus, l'effet de ce facteur n'a été démontré que localement, il est donc intéressant de voir s'il peut être généralisé. La fermeture du milieu a donc été choisie pour tester son influence sur les variables traits précédemment sélectionnées. Cette variable a été transformée en pourcentage afin de compiler les données de chaque relevé où a été retrouvée chaque espèce, pour n'avoir plus qu'un seul indice de fermeture du milieu par espèce. La valeur de cette nouvelle variable –quantitative- étant très sensible au nombre de relevés faits dans chaque type de milieu (fermé, intermédiaire, ouvert), elle a subi une correction (cf. Résultats) afin de supprimer cet effet. Les protocoles différant entre les sources, il a été choisi de travailler sur les présences/absences des espèces et pas sur les densités. Les liens entre les variables –qualitatives- de traits et la variable –quantitative- environnementale ont été testés avec des ANOVA. Ces tests n'ont cependant été réalisés que dans deux sous-échantillons de 39 et 34 espèces faute de temps. La petite taille de ces échantillons nécessite donc de prendre des précautions dans l'interprétation des résultats, car cela ne leur confère qu'une faible force statistique. Ceci permet néanmoins de comparer deux méthodes de choix des espèces via la différence des résultats observés.

### **III. Résultats**

#### **A. Apport et modifications pour les différentes tables de COLTRAIT**

##### 1. Tables « descriptif-relevés »

L'objectif est ici de recueillir un maximum de données issues d'articles sur les espèces de collemboles présentes dans un milieu donné.

##### a) Recherche d'article

La première étape est de sélectionner les articles. Cette sélection répond à plusieurs contraintes générales :

- L'article doit concerner des études de terrain ou éventuellement des expériences menées *in situ*.

- Il ne doit pas être ajouté sur le milieu un intrant qui n'y était pas auparavant ou qui sort de la dynamique culturelle du terrain.
- Un minimum d'information est attendu quant à la description du terrain (au moins ouverture/fermeture du milieu).
- L'identification des collemboles prélevés doit être exhaustive et doit être précisée jusqu'à l'espèce.
- L'article doit être écrit dans une langue relativement familière au lecteur (Français, Anglais, Espagnol).
- La date de publication ne doit pas être antérieure à 1970-60 (l'idéal pas avant 1980) pour des raisons de fiabilité dans les identifications d'espèce. En effet la taxonomie des collemboles et notamment la précision des caractères pour leur identification a fortement évoluée.
- L'échantillonnage doit être fait en Europe.
- L'extraction et l'identification doivent être faites peu de temps après le prélèvement sans conservation vivante dans un microcosme ou autre afin d'éviter l'apparition ou la disparition d'individus.

Un article peut être retrouvé à plusieurs endroits :

ISI Web of Knowledge : les articles peuvent y être téléchargeables en format PDF s'ils ne sont pas trop vieux (>1990-2000).

Dans la bibliothèque du MNHN à Brunoy : un certain nombre de revues y sont disponibles, notamment Pedobiologia, Revue d'écologie et de biologie du sol / European Journal of Soil Biology, Bulletin d'Ecologie etc.

Cette recherche d'articles a été guidée dans un premier temps par une liste de références fournie par JF Ponge, puis poursuivie par une recherche basée sur les publications relatives aux auteurs déjà utilisées et enfin par une recherche par mot clé (collembola) sur ISI Web of Knowledge.

### **J'ai collecté vingt et un articles.**

Une fois l'article sélectionné, les données doivent en être extraites. Il y a trois catégories de données :

- les données descriptives de relevés de terrain
- les données d'observation de collemboles

- les données d'observation de la végétation

Ces trois catégories sont réparties en trois tables. Les définitions des intitulés des champs (définition des variables) de chaque table sont précisées dans un dictionnaire préalablement préparé par Sandrine Salmon. Dans un premier temps, j'ai complété ces définitions, notamment en précisant l'unité des variables quantitatives, ainsi que les différentes modalités que pouvait prendre chaque variable. Cependant les définitions et le nombre de champs sont appelés à varier au fur et à mesure de l'enrichissement des tables, suite à la rencontre dans les articles de nouveaux éléments non synthétisables, ou de nouveaux paramètres paraissant importants ou fréquents ou à l'inverse peu utiles et absents. Ces modifications ont été reportées sur un document à part, au fur et à mesure des observations.

#### b) Modifications de la structure de la base :

J'ai apporté plusieurs modifications à la structure de la base :

Table des relevés de terrain :

Ajout d'un champ « Cmin » et modification du champ « C » en « Cmax » ; ils correspondent au Carbone total en % dans le sol du relevé.

Définition du type d'humus :

Mor : par défaut un terrain tourbeux est considéré comme un humus de type mor (d'après la traduction du mot danois « mor » = tourbe). Un sol décrit comme étant tourbeux dans un article est donc renseigné comme un humus de type Mor dans la table.

Table d'observation de la végétation :

Un champ « strate » pour indiquer la strate végétale dans laquelle se trouve l'espèce a été ajouté avec les modalités suivantes :

T = la plante se trouve dans la strate arborescente

A = la plante se trouve dans la strate arbustive

H = la plante se trouve dans la strate herbacée

NA = espèces n'entrant pas dans cette classification (exemple : plante de type liane)

#### c) Les données descriptives de relevés de terrain

Cette table, nommée « descriptif relevés» regroupe les informations relatives à la description du milieu (codesite) où ont été faits les prélèvements. Ce milieu doit donc être relativement homogène car il sera mis en relation avec les espèces présentes dans les prélèvements lors de l'utilisation de la base de données pour déterminer notamment la gamme de milieux occupée par chaque espèce.

### **Quatre-vingt onze sites ont été ajoutés.**

#### 2. Les données d'observation de collemboles :

La table, nommée « Obscoll », regroupe les espèces de collemboles identifiées dans chaque prélèvement (code-site). Ce code est généralement constitué des trois premières lettres du genre et de l'espèce, cependant cette règle est parfois amenée à être modifiée pour éviter les doublons (exemple : *Anurida granulata* = Anugra et *Anurida granaria* = Anugrar).

Les espèces de collemboles peuvent être nommées sous différents synonymes, notamment à cause des changements de genre de l'espèce.

Pour les vérifications de ce type on peut utiliser le site Fauna Europaea qui répertorie un très grand nombre d'espèces avec leur distribution par pays, leur changement de genre, l'auteur correspondant et la date. Cet outil est très précieux, mais pas toujours à jour, et pas suffisamment complet. Les noms de plusieurs espèces ont donc été rectifiés/validés par Louis Deharveng (OSEB, MNHN) spécialiste de la systématique des collemboles.

J'ai enrichi les champs « nom de l'espèce », « codecoll » et « remarques », de la table « Traits » suite à ces recherches.

Les données sont parfois présentées sous forme de tableau dans les articles. Aussi pour gagner du temps j'ai mis au point la méthode suivante :

#### - Utilisation d'un convertisseur PDF :

J'ai recherché sur internet un programme de conversion de fichier. Ce convertisseur va transformer le tableau du format PDF vers le format Excel par reconnaissance des caractères. Il existe plusieurs programmes de conversion gratuits téléchargeables sur internet. Pour ma part j'ai utilisé : DeskUNPDF

(<http://www.01net.com/telecharger/mac/Bureautique/fiches/101108.html>). Ces programmes sont plus ou moins performants, et tous les tableaux ne peuvent être convertis. Le tableau doit être de préférence entièrement rempli (pas de blanc à la place du zéro, ou de tiret).

### - Tableau converti

Une fois le tableau Excel obtenu, il faut faire une vérification rapide des données afin d'éviter tout problème de mise en forme (colonne décalée ou chiffres manquants...).

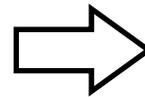
Le tableau Excel peut être ensuite remanié (remplacement des noms de prélèvement de l'article par les codesites, élimination des colonnes ou lignes de données inutiles), afin de ressembler au format de la base.

Figure 5 : Décroisement d'un tableau

### - Décroisement du tableau :

Souvent le tableau répertorie les espèces en fonction du site, informations organisées en tableau croisé (exacte transcription des feuilles de terrain classique). Or la forme des données pour être

	a	b
A	1	2
B	3	4



a	A	1
a	B	3
b	A	2
b	B	4

intégrées dans la table doit être : une colonne pour l'espèce, une colonne pour le site et une colonne pour la valeur relative à l'effectif ; il faut donc décroiser le tableau (fig. 6).

Pour cela j'ai utilisé Decroiser3000 (Sahl, unpublished). Ce programme utilise des formats CSV, aussi le tableau Excel devra être préalablement enregistré sous ce format. Après avoir indiqué un fichier d'accueil du tableau décroisé (créé par le programme et nommé par l'utilisateur) et indiqué si l'on veut prendre en compte les zéros ou pas (cad. si zéro est une information nulle ou si c'est une absence d'information), on lance le programme. Le fichier d'arrivée (format CSV) devra être vérifié. Pour cela un moyen simple est de vérifier qu'il y a le bon nombre de lignes (= nombre de cases pleines du tableau remanié) ou de sommer les cases de la colonne des valeurs et de comparer cette somme à la somme des cases du tableau du fichier Excel remanié.

### -Codage des espèces :

Il faut remplacer les noms d'espèces par leurs codes à l'aide de la table « Traits ». Cette phase peut être faite soit sur le tableau final soit sur le tableau Excel lors du remaniement. Ensuite, il reste à copier ce tableau sur le tableau d'observations des collemboles.

**J'ai ajouté deux mille vingt sept observations à la table « obs-coll ».**

### 3. Les données d'observation de la végétation

En de rares occasions seulement, la végétation du milieu sur lequel a été fait le prélèvement est décrite. Dans ce cas, on répertorie dans la table « Plantes » les espèces trouvées (toujours mises en relation avec le codesite) et leur strate (ex : arborescente).

#### 4. Table « Traits »

L'objectif a été ici, dans un premier temps, de compléter la base en complétant les traits non renseignés pour les espèces déjà présentes dans la table « Traits », et ce, avec différentes clés d'identification de collemboles.

De plus, soixante-douze nouvelles espèces ont été trouvées dans les relevés. Les traits de plusieurs espèces ont été complétés à l'aide des clés de détermination (cf. Bibliographie).

**Tableau 1** présentant les différents champs de la table « Traits », leurs modalités, leur rôle informatif et le niveau de renseignement

Traits/informations	Modalités explorées	Rôle/information	Degrés de renseignement (/510)
Famille		Information taxonomique	Complet (510)
Reproduction	Sexe connu	Reproduction	Peu renseigné (108)
Forme du corps	Elancée, trapue, globuleuse (quantitatif et qualitatif)	Déplacement dans le milieu	Bien renseigné (278)
Coloration	Blanc ou coloré (et détails de la couleur)	Luminosité et couleur du milieu	Bien renseigné (329)
Ecailles	Présence absence	Relation à la luminosité	Presque complet (436)
Taille	Qualitative, quantitative (max et min)	Déplacement dans le milieu	(334) qualitative
Pattes	Longueur par rapport au corps	Déplacement	Bien renseigné (362)
Organes visuels	Présence et nombre de cornéule/ocelle par côté	Organe sensoriel	(326)
Furca	Présence, taille relative (qualitatif)	Déplacement dans le milieu	(335) présence/absence (221) qualitative
Antennes	Taille relative, qualitatif et quantitatif (max min)	Organe sensoriel et déplacement dans le milieu	(242) qualitative (138) quantitative
PAO	Présence et nombre de lobe (max min)	Organe chimio-sensoriel	(248)
Trichobothries	Présence	Organes sensoriel	Bien renseigné (358)
Pseudocelles	Nombre sur l'ensemble du corps	Organe de défense	(89) peu renseigné

a) Structure de la table « Traits » :

Les différents champs sont donnés dans le dictionnaire des données en annexe n°1. La table « Traits » est composée d'un champ spécifiant le nom de l'espèce mais également d'un champ accordé aux remarques quant aux synonymies ou autres observations taxonomiques et un champ à la famille. Elle comporte un champ « codecoll » qui contient un code correspondant à l'espèce de la même ligne. (Rappel) Suivent ensuite toutes les colonnes relatives aux traits avec, couplée à chacune, une colonne de « référence » afin de donner la source de l'information. J'ai complété, pour une partie des 510 espèces que comporte la base, les 25 traits (cf. tableau 1) renseignés dans la base.

b) Modification de la structure de la table « Traits » :

Un champ « longueur corps sur largeur » a été ajouté. Il correspond au ratio de la longueur du corps - de la tête à l'extrémité de l'abdomen - sur sa largeur - mesurée à l'endroit le plus large (milieu) du segment abdominal 3 - . Ce champ est mis en relation avec le champ « forme » et définit numériquement les modalités de cette variable. La relation a été choisie arbitrairement, à partir de dires d'experts d'après des observations de l'allure du corps sur des photos ou des spécimens vivants, de la manière suivante : Ratio longueur/largeur < 3 => Forme= Trapue, Ratio longueur/largeur > 3 => Forme = élancée. La forme globulaire n'est pas prise en compte dans cette relation car cette forme est particulière et propre aux sous-ordres des Symphypleones et des Neelipelones.

La relation entre les champs « long-furca » et « Furca » a été modifiée. Ainsi si le champ « long-furca » = rudimentaire alors le champ « Furca » = non .

c) Création d'une table supplémentaire sur la longueur du corps et du ratio antenne/tête :

A l'usage il est apparu que le dimorphisme sexuel des collemboles et la variabilité intrinsèque au niveau de la taille du corps et du ratio antenne/diagonale de la tête était régulièrement chiffré par les auteurs. Nous avons donc créé une table en complément de la table « Traits » qui pourra par la suite y être insérée. Elle répertorie, si l'information est présente, les différentes tailles trouvées dans les publications (max, min, sexe) pour la taille du corps et le ratio antenne/diagonale de la tête en fonction de l'espèce (nom de l'espèce et codecoll) et de la source (auteur, titre et année de parution de la clé).

#### d) Création d'un tableau de caractères aidant au renseignement de la table

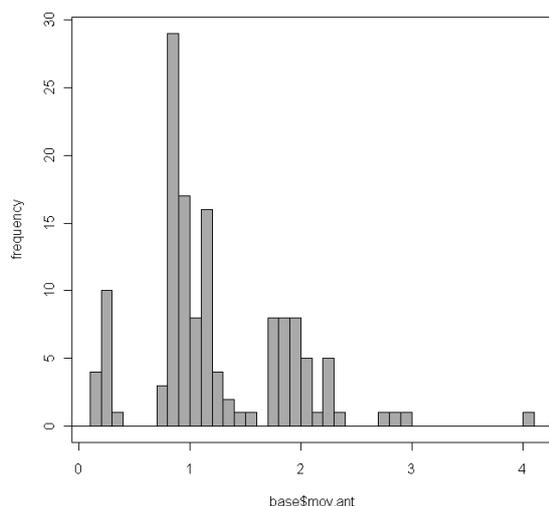
J'ai créé un tableau (schémas en annexe n°4) répertoriant les caractères définis au niveau du sous-ordre, de la famille, voire du genre dans les différentes clés utilisées. Ce tableau évite ainsi des recherches redondantes pour des traits partagés par les espèces d'un même taxon. Ces traits ne sont valables que pour les espèces trouvées dans la/les clé(s) d'identification d'où ont été tirés ces traits, c'est pourquoi chaque trait cité est associé à sa référence.

### **B. Distributions et relations entre les traits morphologiques et physiologiques.**

La distribution des variables ainsi que leurs interprétations ont été analysées sur le jeu de données entier (510 espèces).

Les relations entre les différents traits ont plus ou moins toutes été testées ; cependant, certains champs sont peu renseignés (ex. Pseudocelles, reproduction) ou ont un rôle peu clair (ex. écaille, forme du corps). De plus, un certain nombre de champs (ex. coloration, famille) ont intuitivement un rôle plus ou moins important selon ce qui nous intéresse. On se focalisera donc ici sur les traits qui paraissent les plus pertinents et qui ont été bien renseignés (afin de comparer les traits d'un grand nombre d'espèces).

#### 1. Relations entre les traits quantitatifs

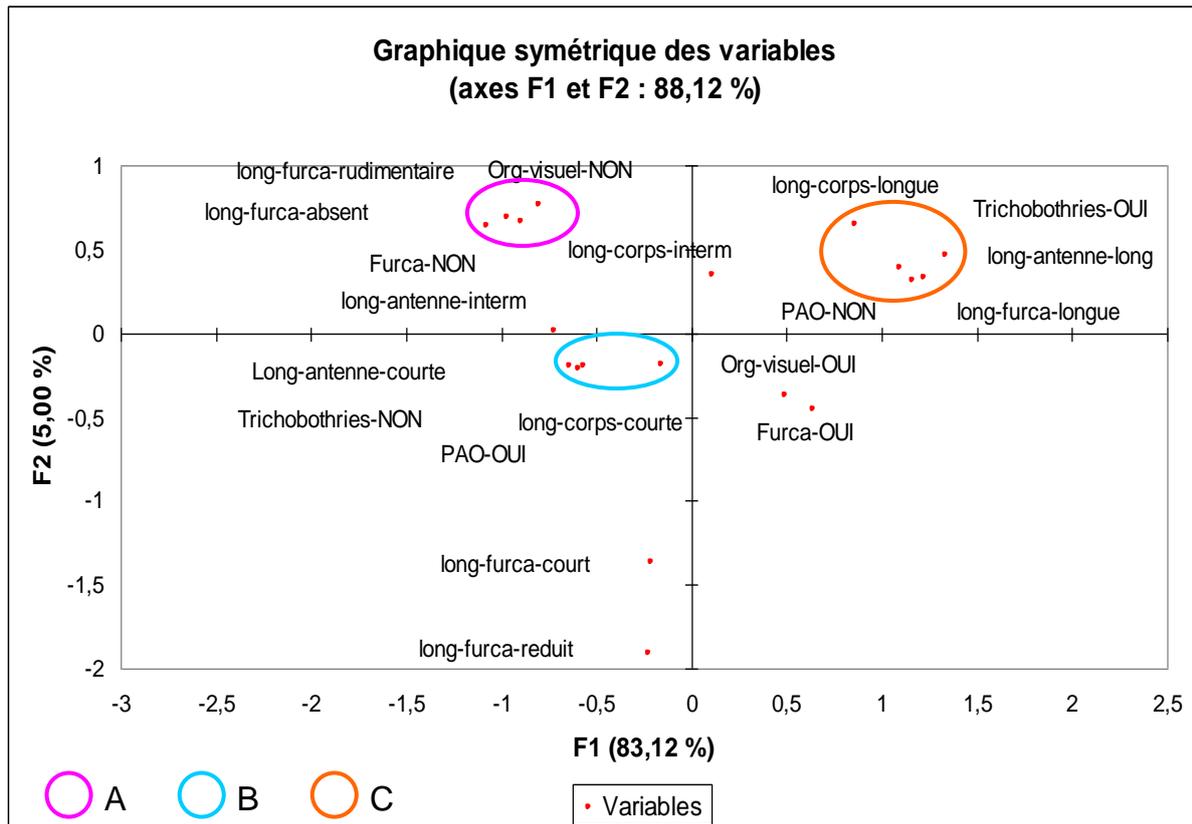


**Figure 6. Distribution de la variable « ratio moyen de la longueur des antennes sur la diagonale de la tête »**

Le ratio antenne/diagonale de la tête est significativement ( $p$ -value = 0,00028) mais légèrement (coef = 0,31) corrélé avec la taille de l'espèce. On remarque également que ce ratio se décompose en patch (3 patch : un après 0, un autour de 1 et un autour de 2). On peut estimer que cette variable est influencée par la famille ( $p$ -value =  $2,2 \cdot 10^{-16}$ ) à laquelle appartient le collembole plutôt que par sa taille ( $p$ -value =  $3,352 \cdot 10^{-8}$ ). Le détail du test se trouve en annexe n°5.

## 2. Relations entre les variables traits qualitatives

Afin de mettre en évidence les différentes relations au sein des modalités de différentes variables traits (cf. plus haut pour le choix de ces variables), une ACM a été réalisée sur les variables qualitatives suivantes : longueur de la furca, longueur relative des antennes, présence/absence de PAO, présence/absence de trichobothries, longueur du corps, organes visuels.



**Figure 7. Analyse par Correspondance Multiple : Graphique symétrique des variables.**  
Cf. texte pour A, B et C.

L'Analyse par Correspondance Multiple (ACM) permet de tester les associations entre les modalités des variables qualitatives choisies. Cette représentation permet d'expliquer 88 % de la variabilité observée. Cette représentation en deux dimensions est donc une bonne image de la variabilité réelle. On observe trois groupes d'associations, plus ou moins fortes, de modalités dans le jeu de données complet (510 espèces) :

**Tableau 2. Groupes d'associations représentés sur la figure 8**

A	une furca absente ou rudimentaire une absence d'organe visuel
B	des antennes de longueur courtes l'absence de trichobothries

	la présence de PAO une longueur de corps courte	Ces corrélations entre traits morphologiques-écophysiologiques vont permettre d'une part de sélectionner les traits qui seront testés pour leur corrélation avec les caractéristiques environnementales, et d'autres ces groupes peuvent correspondre à des écomorphes qui pourront par la suite être mis en relation avec des habitats
C	un corps long la présence de trichobothries une absence de PAO des antennes longues une furca longue	

## **C. Relation entre les traits écologiques et morphologiques : effet de la fermeture du milieu**

### A) Préparation des données :

#### 1) Sélection des espèces

Les analyses ont été faites à partir d'une table compilant les observations de collemboles (table « Obscoll ») et les relevés terrain (table « descriptif-relevés »). Afin d'avoir des espèces dont la gamme de milieux occupés est la mieux décrite et ce, dans un certain nombre de régions, on a tout d'abord procédé à un tri. Le choix s'est porté sur le nombre de sources plutôt que sur le nombre de relevés car le nombre de relevés par sources diffère selon les sources et ne rend pas compte du nombre de milieux différents étudiés. Les espèces sélectionnées (39 espèces) sont celles trouvées dans au moins 15 articles différents.

#### 2) Indice de fermeture de l'habitat et correction

Cet indice est calculé à partir des indications sur le degré de fermeture de l'habitat (milieu ouvert, milieu intermédiaire, milieu fermé, pour plus de détails cf. dictionnaire en annexe n°1) dans lequel l'espèce est trouvée dans chaque relevé. On ne va s'intéresser qu'aux relevés où le milieu est clairement défini (soit ouvert, soit fermé, soit intermédiaire) et on va calculer la part que chaque type de milieu représente dans le total des relevés où une espèce donnée est trouvée. On obtient alors un pourcentage pour chaque type de milieux dont la somme est égale à 100% (les milieux qui ne sont pas clairement définis ne sont pas pris en compte).

Dans la suite du rapport on parlera –pour raccourcir l'énoncé - de fermeture du milieu (ou d'ouverture du milieu) selon qu'il s'agit du pourcentage de milieu fermé par rapport à tous les milieux où est trouvée une espèce donnée, ou du pourcentage de milieux ouverts. Dans le même but, on parlera « du milieu » d'une espèce pour mentionner la somme globale des milieux où cette espèce est trouvée, caractérisé par le pourcentage de milieu fermé/ouvert par rapport à tous les milieux où est trouvée cette espèce.

Cependant, il existe une forte hétérogénéité dans le nombre d'observations au sein des différents types de milieux, le nombre d'observations effectuées en milieu fermé étant beaucoup plus élevé que dans les autres milieux. On peut donc imaginer une correction visant à diminuer cette influence.

**Tableau 3 Pourcentage des types de milieux retrouvés dans les relevés**

	Relevés		Relevé strict
	Nombre	%	%
Milieu fermé strict	740	69	79
Milieu ouvert strict	121	11	13
Milieu intermédiaire strict	78	7	8
Non renseigné	2	0	
Milieu hybride	137	13	
Total	1078	100	100

Effectif recalculé :

$$\text{effF}' = \text{effF} / 0,79$$

eff = effectif

F = milieu fermé

O = milieu ouvert

Pourcentage en conséquence :

$$\% \text{effFcor} = \text{effF}' / (\text{effF}' + \text{effO}' + \text{effI}')$$

I = milieu

intermédiaire

Idem pour %effOcor et %effIcor

Cor = corrigé

Milieu hybride : milieu pas clairement défini, par exemple à la fois qualifié d'ouvert et d'intermédiaire ou d'intermédiaire et de fermé.

Relevé strict : ne prend en compte que les milieux clairement identifiés, donc sans les milieux dits « hybrides ».

Ci-dessous un exemple de l'application de cette correction sur les observations de *Parisotoma notabilis*.

**Tableau 4 Exemple de correction sur les observations d'une espèce**

	Observation de <i>Parisotoma notabilis</i>		
	Nombre	%	milieu strict %
			milieu strict corrigé %

Milieu fermé strict	470	72,0	78,7	33,4
Milieu ouvert strict	84	12,9	14,1	36,3
Milieu intermédiaire strict	43	6,6	7,2	30,2
Non renseigné	1	0,2		
Milieu hybride	55	8,4		
Total	653	100	100	100

On voit que le biais d'échantillonnage a un fort effet sur la part que chaque type de milieu occupe dans la totalité des relevés. Cette correction est appliquée dans le calcul de l'indice de fermeture de l'habitat de chaque espèce.

## B) Relation entre traits morphologiques et indice de fermeture du milieu

### 1. Analyses sur les 39 espèces les mieux représentées dans la base :

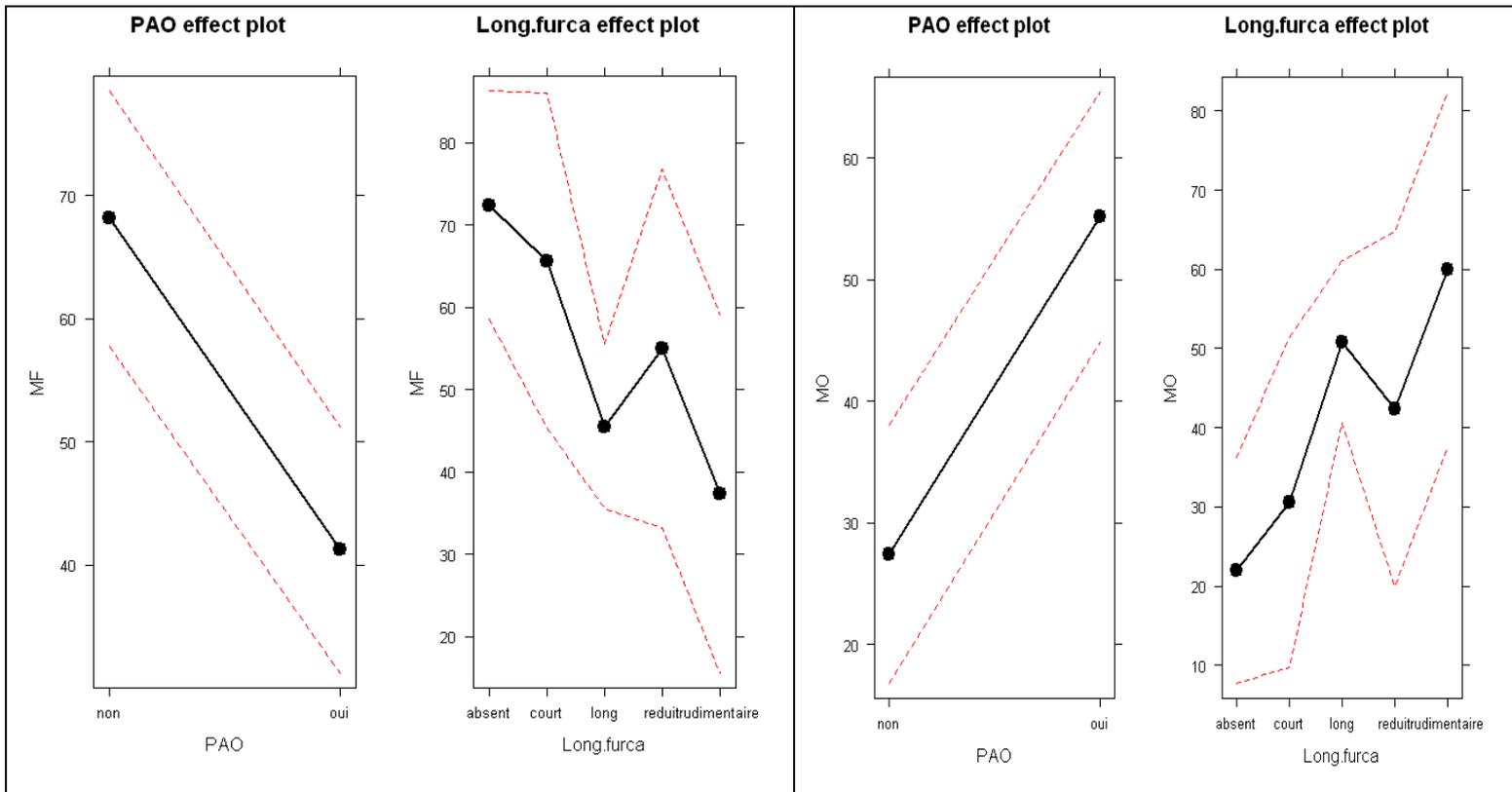
Ce jeu de données a été choisi, comme expliqué précédemment, parce qu'il reposait sur un nombre de sources suffisamment important.

La liste des 39 espèces et leur nombre de sources respectif sont disponibles en annexe n°2.

Des tests (Chi-carré, corrélation de Pearson et de Spearman et ANOVA) pour mettre en évidence des relations entre les différents traits ont été réalisés. On choisit ici de ne présenter que les résultats significatifs. Les détails des tests cités ci-dessous peuvent être vus en annexe n°6. L'indice de fermeture du milieu n'est pas significativement lié avec la longueur des antennes ( $p$ -value= 0,18). En revanche, la présence absence de PAO ( $p$ -value = 0,0019 et  $p$ -value= 0,015) ainsi que la longueur de la furca ( $p$ -value = 0,042) sont liées à l'indice de fermeture du milieu. Ces mêmes facteurs sont également liés avec l'indice d'ouverture du milieu.

**Tableau 5 Effets des facteurs dans les deux ANOVA**

Variation de l'indice de fermeture (ici MF) selon la présence de PAO et la longueur de la furca	Variation de l'indice d'ouverture (ici MO) selon la présence de PAO et la longueur de la furca
---	--



Parmi notre pool d'espèces, celles qui sont largement présentes dans les milieux ouverts possèdent en général une furca longue ou rudimentaire, et sont également caractérisées par la présence de PAO.

**Tableau 6 Effectifs utilisés pour l'analyse**

Longueur Furca	Présence de PAO
Absente : 9	Oui : 18 Non : 20
Courte : 4	
Longue : 18	
Réduite : 3	
Rudimentaire : 3	

Ce résultat va à l'encontre de celui qui a été observé en matière de relation entre longueur de furca et présence de PAO sur la totalité des espèces dans la partie II. Bien que fortement soutenues dans la caractérisation de leur habitat par le nombre de sources dans lesquelles elles apparaissent, ces espèces ne représentent qu'une faible partie du total des espèces répertoriées dans COLTRAIT.

## 2) Jeu de données avec 34 espèces choisies

Face à ces insuffisances et la recherche de la caractérisation des espèces par rapport à leur milieu, un second échantillon a été testé. Cet échantillon (voir en Annexe n°3) est composé à la fois d'espèces apparaissant dans plusieurs sources -supérieur à 5 sources - et à la fois soit

totale absent des milieux ouverts, soit avec une présence dans les milieux ouverts la plus forte possible. Ce choix permet de définir deux groupes, le premier composé d'espèces « spécialistes » du milieu fermé – 20 espèces nouvelles par rapport au jeu de 39 espèces- et le second d'espèces plus « généralistes » -14 espèces reprises dans les 39 espèces précédentes - présentes en part plus ou moins égale dans les deux milieux (ouvert et fermé). Les tests effectués (voir en annexe n°7 pour le détail) montrent que dans l'échantillon, la présence de PAO (p-value = 0,021) est plus fréquente pour les espèces occupant autant les milieux ouverts que fermés, alors que cette fois ni la longueur des antennes (p-value = 0,36) ni la longueur de la furca (p-value = 0,59) ne sont différentes en fonction de l'indice de fermeture du milieu. De plus l'indice de fermeture du milieu est négativement corrélé (p-value = 0,014 ; Coef = -0,58) au nombre de lobes du PAO.

## IV. Discussion

Un point important de l'alimentation de la base COLTRAIT est le choix des articles. Ce choix doit être défini selon des critères précis sans toutefois être trop restrictifs. Précis car les données qui en sont tirées doivent pouvoir être utilisées pour vérifier des hypothèses et tester des modèles, elles doivent donc comprendre un minimum d'informations sur les conditions d'obtention des résultats (protocole, localisation, date, milieu). Sans être trop restrictif car le but reste de rassembler le maximum d'informations pour pouvoir faire des tests avec un maximum de robustesse. Les indications guidant le choix des articles ont été choisies afin d'optimiser ce compromis entre précision et nombre d'articles. Par exemple, le choix important de la région de prospection (l'Europe) doit se faire à la fois en tenant compte de la relative homogénéité en espèces du milieu et du maximum de données disponibles provenant de la région.

La structure de la base (champs et modalités) a été améliorée durant ce travail. Des champs ont été ajoutés et des modalités précisées ou ajoutées. La structure de la base va en effet être à l'image de l'information disponible. Or cette information n'est pas toujours disponible sous une forme adaptée à la base. Les articles d'où les informations sont tirées n'ont initialement pas pour objectif de nourrir la base. De plus la base doit par la suite proposer des données dans une forme telle qu'elles peuvent être utilisées statistiquement. Il faut donc penser à la fois à offrir un maximum de renseignements tout en pensant à normaliser ces renseignements à travers le choix de catégories. L'accumulation de données amène donc à revoir certains points de sa structure.

La base offre de nombreuses possibilités de test. Une sélection a été faite autour d'un thème : la relation des traits des espèces à l'indice de fermeture et/ou d'ouverture de l'ensemble des milieux où elles sont trouvées. Ce thème offre encore un certain nombre de possibilités, aussi un grand nombre de tests et d'observations de variables, qui ont été faits pour sonder les points intéressants du jeu de données, ont été réalisés mais n'ont pas été montrés ici. Le choix a été de se pencher uniquement sur trois traits : la longueur des antennes, la longueur de la furca et la présence/absence de PAO.

Les analyses sur les relations entre traits montrent plusieurs choses :

➤ Le trait « longueur relative des antennes » à une distribution en 3 patches. Le graphique montre que même s'il y a une corrélation entre taille et longueur relative des antennes, celle-ci est faible et on observe une grande variabilité de la taille pour un même patch. Ces constatations amènent à se poser la question de l'influence de la taxonomie sur les traits morphologiques.

➤ Sur la totalité des espèces répertoriées dans la base, la longueur des antennes, la longueur de la furca et la présence/absence de PAO sont corrélés entre eux. Le PAO est un chimio-récepteur et est un organe sensoriel chez les collemboles. Cependant son rôle exact est encore mal connu. La présence ou l'absence de PAO semble associé aux deux autres traits selon un pattern particulier :

Présence de PAO, antennes courtes

Absence de PAO, furca longue, antennes longues

L'utilisation de sous-échantillons du jeu de données complet afin d'étudier la relation des traits au milieu a nécessité une préparation des données. La correction réalisée sur les données à partir des pourcentages des différents types de milieux dans les relevés, a permis de corriger un des déséquilibres possibles présent dans la collection des données. Il peut exister d'autres biais à corriger, notamment l'effet de l'homogénéité de l'origine des sources d'observation. En effet, les articles sont réalisés à partir de prélèvements faits en Europe. Cependant il existe des disparités dans les milieux et les milieux ne sont pas tous représentés de manière égale (nombre d'observations). Cela crée donc un biais dans la représentation des communautés de collemboles.

Suite à cette correction les relations entre les traits et le milieu ont été testés dans deux jeux de données. Lorsque l'on sélectionne un pool d'espèces sur la base de leur représentativité dans

la base (nombre d'articles), les espèces retrouvées essentiellement dans des milieux fermés sont caractérisées par une absence de PAO et une furca « courte » ou absente (la catégorie « rudimentaire » allant dans le sens contraire). A l'inverse, la longueur relative des antennes ne semble pas entrer en jeu dans la caractérisation des espèces en fonction de la fermeture de leur milieu. Les autres traits morphologiques testés (cf. tableau 1) ne sont pas significativement liés à la fermeture du milieu dans cet échantillon.

Le second jeu de données, sélectionné selon la préférence pour la fermeture du milieu a en premier lieu été analysé afin de déterminer s'il existe des espèces spécialistes de milieux forestiers et spécialistes de milieux ouverts. La notion de spécialisation est reliée à la notion de niche écologique. Elle est souvent définie différemment selon le contexte (Devictor et al. 2010). Une espèce spécialiste aura une niche écologique plus réduite par rapport à une espèce généraliste pour un ensemble de paramètres. Cette notion est donc relative à ce que l'on compare. Ici on compare les espèces en fonction de la fermeture/ouverture du milieu.

D'autre part il existe plusieurs moyens de déterminer le caractère spécialiste ou généraliste d'une espèce. On peut choisir un caractère continu en établissant un indice de spécialisation : par exemple, le taux de présence/absence dans un jeu de classes d'habitat (Devictor et al, 2008) ; une variation dans la densité d'individu entre ces classes d'habitat (Julliard et al, 2006) ; ou bien définir une limite stricte en choisissant, à partir d'un indice, un seuil à partir duquel l'espèce est spécialiste (ou généraliste).

Dans nos analyses, plus une espèce aura des indices équilibrés entre eux plus elle sera généraliste et à l'inverse plus les indices seront déséquilibrés, plus elle sera spécialiste. Ici on choisit une spécialisation stricte, c'est-à-dire que l'espèce ne doit pas être trouvée dans un milieu ouvert pour être définie comme spécialiste du milieu fermé ; le milieu intermédiaire est laissé de côté.

Nous n'avons pas trouvé d'espèces spécialistes de milieu ouvert, mais ceci est peut-être dû au fait que les études répertoriées dans COLTRAIT proviennent majoritairement de milieu forestier. Ce point peut constituer un biais, mais son importance est à nuancer, puisque c'est parce que les collemboles sont plus abondants en milieux forestiers qu'ils y sont le plus étudiés. Afin de préciser l'étude des spécialistes/généralistes une comparaison des espèces sur la base de leurs densités et non pas seulement de la présence/absence des espèces devra être réalisée. Ainsi la caractérisation des communautés se ferait par rapport aux variations observées dans la densité d'individus de ces espèces.

Lorsque l'on sélectionne les espèces selon leur préférence/indifférence pour la fermeture du milieu, seule la présence/absence de PAO semble caractériser les espèces en fonction de la fermeture de leur milieu. Les autres traits morphologiques de la base ne sont pas significativement liés à la fermeture du milieu dans cet échantillon. L'absence de PAO caractérise des espèces trouvées majoritairement dans des milieux fermés. L'échantillon est plus représentatif des parts de chaque forme de traits et des associations de traits - absence de PAO, longue furca – du jeu de données complet que de l'échantillon des 39 espèces.

La place importante du PAO dans la caractérisation des communautés en fonction de l'ouverture/fermeture du milieu est confirmée par l'analyse de ce second jeu de données. De plus, cette relation est renforcée par la corrélation entre le nombre de lobes du PAO et l'ouverture du milieu. Aucune étude n'avait, auparavant montré une telle relation du PAO avec l'ouverture du milieu. Le milieu ouvert étant un milieu plus changeant (hygrométrie, température etc.), le PAO pourrait alors être un atout dans la perception de ces changements défavorable/favorable et permettrait à l'individu de se déplacer en conséquence.

**Tableau 7. Résumé des Résultats**

	Jeu de données traité	Jeu de 39 espèces	Jeu de 34 espèces choisies
Caractère du jeu de données	Complet	Les plus communes	Deux groupes associés à deux tendances de milieu différentes
Antenne	Le caractère long est associé à une absence de PAO	Pas de lien avec la fermeture du milieu	Pas de lien avec la fermeture du milieu
Fruca	Le caractère long est associé à une absence de PAO	Lié au milieu Longueur associé à un milieu majoritairement fermé	Non lié – mise en évidence de l'effet d'échantillonnage
PAO	Associé à des antennes longues et une furca longue	Absence de PAO lié à un milieu majoritairement fermé	Absence de PAO lié à un milieu majoritairement fermé

Les résultats à priori contradictoires (tableau 7) posent, dans un premier temps, la question du choix du jeu de données sur lequel va porter l'analyse par rapport à sa représentativité du jeu de données de traits morphologiques complet (510 espèces).

Deux questions se posent :

- Faut-il choisir une représentativité par rapport à la diversité des traits écologiques (comme la préférence/indifférence pour la fermeture du milieu)?
- Ou faut-il choisir une représentativité par rapport à la fréquence des espèces retrouvées qui forment une caractéristique stable et importante des populations observées ?

Ce choix est déterminé par la question que l'on se pose. Dans le cas d'un questionnaire comme le notre, il faudrait à la fois sélectionner en fonction du trait écologique et également augmenter le nombre d'espèces analysées, notamment pour les espèces généralistes (uniquement 14 espèces pour 20 espèces de spécialistes).

D'autre part, la contradiction des résultats dans le sens des corrélations entre les traits morphologiques, selon le jeu de données considéré pourrait s'expliquer par l'interaction de plusieurs facteurs qui s'exprimeraient différemment selon les espèces sélectionnées. Il s'agit en particulier du facteur « profondeur » et « ouverture du milieu ». En effet, les adaptations morphologiques à la profondeur (réduction de la furca, des antennes, absence d'organe visuels...) sont bien connues chez les collemboles, bien qu'elles n'aient jamais fait l'objet d'analyse à grande échelle, sur un grand nombre d'espèces (Hopkin, 2007). Les caractères supposés permettre l'adaptation au milieu ouvert c'est-à-dire un déplacement rapide et des organes sensoriels nombreux pour capter les variations environnementales plus importantes (furca longue, antennes longues, PAO) sont aussi ceux qui conditionnent un mode de vie épigé. Par conséquent la sélection des espèces selon l'ouverture/fermeture du milieu pourrait permettre de révéler les caractères liés à l'ouverture/fermeture du milieu, qui sont masqués par la relation avec la profondeur dans le pool d'espèces total. De plus, lorsqu'on sélectionne les espèces les mieux représentées, on a de fortes chances de sélectionner des espèces « généralistes », ce qui correspond aussi finalement à une sélection selon un critère écologique.

De plus, le manque d'information de certains traits par rapport à d'autres et surtout en fonction de la famille introduit des biais importants dans l'interprétation de leur distribution. Ce défaut se corrigera sans doute au fur et à mesure de l'alimentation de la base, mais on peut y pourvoir dès maintenant en choisissant des espèces qui ont un niveau de renseignement identique pour les traits comparés.

Dans un autre ordre d'idée il serait intéressant de faire une étude de la distribution géographique des espèces et des sources afin de cibler d'éventuels biais qui pourraient influencer la perception de la relation milieu/traits.

## V. Conclusion

La base de données COLTRAIT permet de gérer un grand nombre d'informations de natures très différentes (traits morphologiques, caractéristiques du milieu, effectifs). Cette caractéristique est appréciable quand le but est de mener une étude à grande échelle comme celle de COLTRAIT.

Les comparaisons faites entre les traits choisis (PAO, longueur de la furca, longueur des antennes) en fonction des différents milieux ont permis de se rendre compte à l'usage des améliorations à apporter à la méthode d'analyse et de sélection des données. De plus, quelques résultats ont permis de réfléchir à un certain nombre de pistes à explorer pour la suite, notamment quant au rôle de l'organe post-antennaire dans la fitness de l'espèce et les avantages qu'il apporte dans un milieu ouvert.

## VI. Bibliographie

1. Bååth, E., Berg, B., Lohm, U., Lundgren, B., Lundkvist, H., Rosswall, T., Söderström, B., Wirén, A., 1980. Effects of experimental acidification and liming on soil organisms and decomposition in a Scots pine forest. *Pedobiologia* 20, 85-100.
2. Bonnet, L., Cassagnau, P., Deharveng, L., 1976. Un exemple de rupture de l'équilibre biocénotique par déboisement: les peuplements de collemboles édaphiques du Piau d'Engaly (Hautes-Pyrénées). *Revue d'Écologie et de Biologie du Sol* 13, 337-351.
3. Bretfeld G., 1999, Synopses on Palaearctic Collembola Volume 2 : Symphypleona.
4. Bruckner, G. Barth and M. Scheibengraf, 2000, Composite sampling enhances the confidence of soil microarthropod abundance and species richness estimates. *Pedobiologia* 44, 63-74.
5. Cassagne, N., Bal-Serin, M.C., Gers, C., Gauquelin, T., 2004. Changes in humus properties and collembolan communities following the replanting of beech forests with spruce. *Pedobiologia* 48, 267-276.
6. Chahartaghia M., et al, 2006, Sex ratio and mode of reproduction in Collembola of an oak beech forest. *Pedobiologia* 50, 331-340.
7. D'Haese CA, 2003, Homology and morphology in Poduromorpha (Hexapoda, Collembola) *Eur. J. Entomol.* 101, 385–407.
8. Devictor V. et al, 2008, Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation, *Oikos* 117: 507-514.
9. Devictor V. et al, 2010, Defining and measuring ecological specialization *Journal of Applied Ecology* 47 : 15–25.
10. Dunger, W., Schulz, H.J., Zimdars, B., Hohberg, K., 2004. Changes in collembolan species composition in eastern Germany mine sites over fifty years of primary succession. *Pedobiologia* 48, 503-517.
11. Fjellberg A., 1998, The Collembola of Fennoscandia and Denmark, volume 35, Part I : Poduromorpha.
12. Fjellberg A., 2007, The Collembola of Fennoscandia and Denmark, volume 42, part II : Entomobryomorpha and Symphypleona. BRILL.

13. Fratello, B., Bertolani, R., Sabatini, M.A., Mola, L., Rassa, M.A., 1985. Effects of atrazine on soil microarthropods in experimental maize fields. *Pedobiologia* 28, 161-168.
14. Gachet S. et al, 2005, BASECO: a floristic and ecological database of Mediterranean French flora, *Biodiversity and Conservation* 14: 1023–1034.
15. Hopkin SP, 2007, A key to the Collembola (Springtails) of Britain and Ireland. FSC Publication.
16. Jordana R., et al 1997, Fauna Iberica Volume 8 : Collembola Poduromorpha.
17. Julliard R. et al, 2006, Spatial segregation of specialists and generalists in bird communities, *Ecology Letters* 9: 1237–1244.
18. Kováč, L., Kostúrova, N., Miklisová, D., 2005. Comparison of collembolan assemblages (Hexapoda, Collembola) of thermophilous oak woods and *Pinus nigra* plantations in the Slovak karst (Slovakia). *Pedobiologia* 49, 29-40.
19. Lauga-Reyrel, F., Lauga, J., 1995. Collembola of cold Pyrenean habitats. *European Journal of Soil Biology* 31, 217-229.
20. Lauga-Reyrela F., Marc Deconchat, 2000, Diversity within the Collembola community in fragmented coppice forests in south-western France. *European Journal of Soil Biology* 35(4), 177-187.
21. Liiri, M., Haimi, J., Setälä, H., 2002. Community composition of soil microarthropods of acid forest sites as affected by wood ash application. *Pedobiologia* 46, 108-124.
22. Lopes, C.M., da Gama, M.M., 1994. The effect of fire on collembolan populations of Mata da Margaraça (Portugal). *European Journal of Soil Biology* 30, 133-141.
23. Martin Holmstrup, Kristine Maraldo, Paul Henning Krogh, 2007, Combined effect of copper and prolonged summer drought on soil Microarthropods in the field. *Environmental Pollution* 146, 525-533.
24. Nathalie Cassagne, Thierry Gauquelin, Marie-Claude Bal-Serin, Charles Gers, 2006, Endemic Collembola, privileged bioindicators offorest management. *Pedobiologia* 50, 127-134.
25. Pitzalis M. et al , 2010, Co-occurrence analyses show that non-random community structure is disrupted by fire in two groups of soil arthropods (Isopoda Oniscidea and Collembola). *Acta Oecologica* 26, 100-106.

26. Pomorski RJ, 1998, Onychiurinae of Poland ( Collembola : Onychiuridae)
27. Potapow M., 2001, Synopses on Palaearctic Collembola, volume 3 : Isotomidae.
28. Querner P., A. Bruckner, 2010, Combining pitfall traps and soil samples to collect Collembola for site scale biodiversity assessments. *Applied Soil Ecology* 45, 293-297.
29. Rebecchi, L., Sabatini, M.A., Cappi, C., Grazioso, P., Vicari, A., Dinelli, G., Bertolani, R., 2000. Effects of a sulfonylurea herbicide on soil microarthropods. *Biology and Fertility of Soils* 30, 312-317.
30. Russell DJ., G. Alberti, 1998, Effects of long-term, geogenic heavy metal contamination on soil organic matter and microarthropod communities, in particular Collembola. *Applied Soil Ecology* 9, 483-488.
31. Servane Gillet, Jean-François Ponge, 2005. Species assemblages and diets of Collembola in the organic matter accumulated over an old tar deposit,. *European Journal of Soil Biology* 41 (2005) 39–44.
32. Shaw, P.J.A., Usher, M.B., 1996. Edaphic Collembola of lodgepole pine *Pinus contorta* plantations in Cumbria, UK. *European Journal of Soil Biology* 32, 89-97.
33. Syrek, D., Weiner, W.M., Wojtylak, M., Olszowska, G., Kwapis, Z., 2006. Species abundance distribution of collembolan communities in forest soils polluted with heavy metals. *Applied Soil Ecology* 31, 239-250.
34. Thibaud JM, HJ. Schulz, MM da Gama Assalino, 2004, Synopses on Palaearctic Collembola, Volume 4 : Hypogastruridae.
35. Violle C., 2007, Let the concept of trait be functional!, *Oikos* 116: 882-892.
36. Volkmar Wolters, 1998, Long-term dynamics of a collembolan community. *Applied Soil Ecology* 9, 221-227.
37. Weiher E. et al, 1995, Assembly Rules, Null Models, and Trait Dispersion: New Questions from Old Patterns, *Oikos*, 74 : 159-164

## Annexe 1 : Description des tables et champs de COLTRAIT

### **Table « traits »**

#### **« Codecoll » : clé primaire**

Tables traits morphologiques, écophysiologicals et d'histoire de vie.

Champ « Remarques »

« nomcoll »

« codecoll »

« Famille »

Les références des informations renseignées sont à indiquer dans chaque colonne « Ref-nom-de-la-colonne-renseignée ». Les informations sont divisées en groupes.

« Reproduction »

« Mâle-femelle »

« Forme »

« Rapport longueur/largeur » :

« Coloration »

« Couleur »

« Long-corps »

« Longtotmin »

« Longtotmax »

« Ecailles »

« Furca »

« Long-furca »

« Lg-pattes/Lg-corps »

« Pseudoce tot max »

« Pseudoce tot min »

« Trichobothries »

« Long-antennes/tête max »

« Long-antennes/tête min »

« Long-antennes »

« Org-visuel »

« Nb-cornéulesmin/coté »

« Nb-cornéulesmax/coté »

« PAO »

« Nb min lobes/PAO »

« Nb max lobes/PAO »

### **Table « Obs-collemboles »**

Champ « codesite »

Champ « codecoll »

Champ « obs »

Champ « nature-obs »

### **Table « végétation »**

Espèces d'arbres, de plantes et de mousses :

« nom »

« obs »

« Strate »

## Table « descriptif-relevés »

« Codesite »	« Hz A »	« pHmax »
« Surface »	« Hz B »	« Landes-Buisson »
« Volume »	« HzE »	« Bois »
« AbdT »	« Hz-org »	« Artificial forest »
« Richesse spécifique »	« Hz-org-min »	« Forêt-feuillus »:
« Richesse relative »	« Prof0 »	« Forêt-conif »
« Mois-saison »	« Prof1 »	« Forêt mixte »
« Année » .	« Prof2 »	« Coupe-à-blanc »
« Commune »	« Prof3 »	« Clairière »
« Pays »	« Prof4 »	« Haie »
« Coord ».	« SYN-Sol »	« Prairie «
« Altmin »	« Sol-cultivé «	« Prairie-fauche »
« Altmax »	« Mull »	« Pâturage »
« Pluviom-min »	« Eumull»:	« Jachères »
« Pluviom-max »	«Mesomull» :	« Champs-cultivés »
« Temperature-min »	«Oligomull» :	« Bord-eau «
« Temperature-max »	«Dysmull»	« Chemin »
« AgeSitemin »	«Amphimull»	« Littoral »
« AgeSitemax »	« Moder » :	« Marais salés »
« Source »	« Eumoder »	« Sable-dunes »
« Roche »	« Dysmoder »	« Milieu-ouv »
« Part-aer-plantes »	« Hemimoder »	« Milieu- fer »
« Mousses-sol »	« Amphimoder »	« SYN-Milieu-ouv »
« Raci-plantes »	« Mull-moder »	« SYN-Milieu-fer »
« Tronc-arbre »	« Mor » :	« SYN-Milieu-inter »
« Canope »	« SYN-Mull »	« SYN_edaph »
« Bois-au-sol »	« SYN-Moder »	« SYN_hemiedaph »
« Souche-arbre »	« SYN-Mor »	« SYN_epige »
« Humus-bois »	« Hydromorphie »	« SYN-super-épigée »
« Deject-mammif »	« C »	« SYN-hyper-épigée »
« Deject-lombrics »	« Nmin »	« SYN-Hydrophilie »
« Debris-veg »	« Nmax »	
« Dech-pol »	« NH4 »	
« Mycelium »	« NO3 »	
« Dans-eau »	« C/Nmin »	
« Hz L »	« C/Nmax »	
« Hz F »	« MOT »	
« Hz H »	« pHmin »	

**SYN-Milieu-ouv** : 1 = colonne Milieu-ouv + Prairie + Prairie-fauche +Pâturage + Jachères + Champs-cultivés + Marais salés+ Sable-dunes

**SYN-Milieu-fer** : Somme Milieu-fer + Bois+ forêt-feuillus +forêt-conif + forêt mixte + artificial-forest

**SYN-Milieu-inter** = colonnes Milieu-inter + Landes-Buissons + haie + Clairière + coupe à blanc

## Annexe 2 : Jeu de données 39 espèces (>15 sources)

Table des noms d'espèces utilisées :

nomcoll	codecoll
Parisotoma notabilis (Schäffer, 1896)	Parnot
Isotomiella minor (Schäffer, 1896)	Isomin
Folsomia quadrioculata (Tullberg, 1871)	Folqua
Megalothorax minimus Willem, 1900	Megmin
Willemia anophthalma Borner, 1901	Wilano
Neanura muscorum (Templeton, 1835)	Neamus
Micranurida pygmaea Borner, 1901	Micpyg
Sphaeridia pumilis (Krausbauer, 1898)	Sphpum
Paratullbergia callipygos (Borner, 1902)	Parcal
Lepidocyrtus lignorum (Fabricius, 1793)	Leplig
Friesea mirabilis (Tullberg, 1871)	Frimir
Entomobrya nivalis (Linnaeus, 1758)	Entniv
Sminthurinus aureus (Lubbock, 1862)	Smiaur
Pseudosinella alba (Packard, 1873)	Psealb
Mesaphorura macrochaeta Rusek, 1976	Mesmac
Ceratophysella armata (Nicolet, 1841)	Cerarm
Ceratophysella denticulata (Bagnall, 1941)	Cerden
Folsomia manolachei Bagnall, 1939	Folman
Isotoma viridis Bourlet, 1839	Isovir
Mesaphorura hylophila Rusek, 1982	Meshyl
Mesaphorura krausbaueri Borner, 1901	Meskra
Micraphorura absoloni (Borner, 1901)	Micabs
Protaphorura armata (Tullberg 1869)	Proarm
Pseudachorutes parvulus Borner, 1901	Psepar
Tomocerus minor (Lubbock, 1862)	Tommin
Allacma fusca (Linnaeus, 1858)	Allfus
Brachystomella parvula (Schaffer, 1896)	Brapar
Isotomurus palustris (Müller, 1776)	Isopal
Pogonognathellus flavescens (Tullberg, 1871)	Pogfla
Xenylla tullbergi Borner, 1903	Xentul
Dicyrtoma fusca (Lubbock, 1873)	Dicfus
Dicyrtomina minuta (O. Fabricius, 1783)	Dicmin
Friesea truncata Cassagnau, 1958	Fritru
Heteromurus major (Moniez, 1889)	Hetmaj
Heteromurus nitidus (Templeton, 1835)	Hetnit
Mesaphorura yosii (Rusek, 1967)	Mesyos
Orchesella cincta (Linnaeus 1758)	Orccin
Pseudisotoma sensibilis (Tullberg, 1876)	Psesen
Willemia denisi Mills, 1932	Wilden

## Annexe 3 : Jeu de données 34 espèces choisies

Table des noms d'espèces utilisées :

nomcoll	codecoll
Allacma fusca (Linnaeus, 1858)	Allfus
Allacma gallica (Carl, 1899)	Allgal
Anurida granulata Agrell, 1943	Anugra
Anurophorus laricis Nicolet, 1842	Anular
Anurophorus septentrionalis Palissa, 1966	Anusep
Ceratophysella tuberculata Cassagnau, 1959	Certub
Deutonura deficiens Deharveng, 1979	Deudef
Entomobrya albocincta (Templeton, 1835)	Entalb
Entomobrya corticalis (Nicolet, 1842)	Entcor
Gisinianus flammeolus (Gisin, 1957)	Gisfla
Kalaphorura burmeisteri (Lubbock, 1873)	Kalbur
Mesaphorura betschi Rusek, 1979	Mesbet
Microgastrura sensiliata Jordana, 1981	Micsen
Monobella grassei (Denis, 1923)	Mongragr
Orchesella bifasciata (Bourlet 1839)	Orcbif
Pseudisotoma monochaeta (Kos, 1942)	Psemon
Tomocerus vulgaris (Tullberg, 1871)	Tomvul
Vertagopus cinereus (Nicolet, 1841)	Vercin
Xenyllodes armatus Axelson, 1903	Xenarm
Xenylla tullbergi Borner, 1903	Xentul
Folsomia quadrioculata (Tullberg, 1871)	Folqua
Sphaeridia pumilis (Krausbauer, 1898)	Sphpum
Friesea mirabilis (Tullberg, 1871)	Frimir
Sminthurinus aureus (Lubbock, 1862)	Smiaur
Pseudosinella alba (Packard, 1873)	Psealb
Ceratophysella armata (Nicolet, 1841)	Cerarm
Ceratophysella denticulata (Bagnall, 1941)	Cerden
Isotoma viridis Bourlet, 1839	Isovir
Mesaphorura krausbaueri Borner, 1901	Meskra
Protaphorura armata (Tullberg 1869)	Proarm
Brachystomella parvula (Schaffer, 1896)	Brapar
Isotomurus palustris (Müller, 1776)	Isopal
Dicyrtomina minuta (O. Fabricius, 1783)	Dicmin
Heteromurus nitidus (Templeton, 1835)	Hetnit

**Annexe 4 : Schéma de la structure du tableau de traits  
aidant au renseignement de la table**

Sous-Ordre	Caractères du sous ordres		Famille	Caractères défini à partir de la famille	
	références	particularité		références	particularité
Entomobryomorphes					
Neelipleones					
Poduromorphes					
Symphypleones					

## Annexe 5 : Relation entre les traits quantitatifs

Test de corrélation de Spearman bilatéral :

p-value = 0.0002787

Coefficient de corrélation estimé : cor=0.3080194

ANOVA à deux facteurs				
Modèle :				
Log(moy ratio antenne) = [Famille] + [Longueur corps] + [Famille] : [longueur corps]				
Résultats :				
	Sum Sq	Df	F value	Pr(>F)
Famille	27.5259	16	36.0264	< 2.2e-16 ***
longue.corps	1.9329	2	20.2388	3.352e-08 ***
Famille:longue.corps	1.0913	9	2.5392	0.011 *
Residuals	5.2051	109		
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Normalité des résidus (Shapiro-Wilk) : p-value = 0,1164				

- ⇒ les résidus ont une distribution normale
- ⇒ Le modèle est validé

### Annexe 6 : Analyses sur les 39 espèces les mieux représentées dans la base

Modèle : 39 espèces					
MF= [PAO] + [longueur antenne] + [longueur antenne] : [PAO]					
	Sum Sq	Df	F value	Pr(>F)	Shapiro-wilk : p-value = 0.4427 ⇒ normalité des résidus ⇒ Modèle validé
PAO	2627.5	1	6.6028	0.01504 *	
Long.antenne	1431.6	2	1.7988	0.18184	
PAO:Long.antenne	204.9	1	0.5150	0.47819	
Residuals	12733.8	32			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
MF= [PAO] + [longueur furca] + [longueur furca] : [PAO]					
	Sum Sq	Df	F value	Pr(>F)	Shapiro-Wilk : p-value = 0.7728 ⇒ normalité des résidus ⇒ Modèle validé
Long.furca	3643.4	4	2.8657	0.042301 *	
PAO	3764.1	1	11.8425	0.001899 **	
Long.furca:PAO	1824.8	3	1.9137	0.151171	
Residuals	8581.8	27			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

## Annexe 7 : Analyses faites sur le jeu de données de 34 espèces choisies

Modèle : 34 espèces					
MF= [PAO] + [longueur antenne] + [longueur antenne] : [PAO]					
	Sum Sq	Df	F value	Pr(>F)	Shapiro-Wilk : p-value = 0.0171 ⇒ Résidus non normaux ⇒ Modèle non validé
PAO	3552.9	1	3.6297	0.06933 .	
Long.antenne	849.2	1	0.8676	0.36131	
PAO:Long.antenne	1486.3	1	1.5184	0.23031	
Residuals	22513.2	23			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
MF= [PAO] + [longueur furca] + [longueur furca] : [PAO]					
	Sum Sq	Df	F value	Pr(>F)	Shapiro-wilk : p-value = 0.05866 ⇒ normalité des résidus ⇒ Modèle validé
Long.furca	3066.5	4	0.7163	0.58980	
PAO	6563.2	1	6.1320	0.02145 *	
Long.furca:PAO	238.2	2	0.1113	0.89521	
Residuals	23547.2	22			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

**Annexe 7 suite : Analyses faites sur le jeu de données de 34 espèces choisies**

Test pour le milieu fermé	Résultats																						
<p>Test de Mann-Whitney unilatérale</p> <p>H0 : Les espèces sans PAO (15) ont un indice de fermeture moyen de leur milieu égale à celui ayant un PAO (18)</p> <p>H1 : Les espèces sans PAO ont un indice de fermeture moyen de leur milieu supérieur à celui ayant un PAO</p>	<p style="text-align: center;">p-value = 0,011</p> <p style="text-align: center;">⇒ H0 est rejetée</p> <p style="text-align: center;">Moyenne des espèces sans PAO : 68.06 Moyenne des espèces avec PAO : 42.29</p> <p style="text-align: center;">Ces moyennes sont significativement différentes.</p>																						
Test du Chideux entre PAO et longueur de la furca	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th colspan="2">PAO</th> </tr> <tr> <th>Long.furca</th> <th>non</th> <th>oui</th> </tr> </thead> <tbody> <tr> <td>absent</td> <td>0</td> <td>4</td> </tr> <tr> <td>court</td> <td>0</td> <td>3</td> </tr> <tr> <td>long</td> <td>10</td> <td>2</td> </tr> <tr> <td>reduit</td> <td>1</td> <td>4</td> </tr> <tr> <td>rudimentaire</td> <td>3</td> <td>3</td> </tr> </tbody> </table>		PAO		Long.furca	non	oui	absent	0	4	court	0	3	long	10	2	reduit	1	4	rudimentaire	3	3	<p>p-value = 0.007098</p> <p>Condition d'application : 8 fréquences attendues &lt;5</p> <p>⇒ test non validé</p>
	PAO																						
Long.furca	non	oui																					
absent	0	4																					
court	0	3																					
long	10	2																					
reduit	1	4																					
rudimentaire	3	3																					
Test de Corrélation de Spearman Unilatérale cor<0 entre la	<p>Entre deux variables quantitatives non normale (n&gt;30)</p> <p>p-value = 0.01424</p>																						

fermeture du milieu et le nombre de lobe du PAO	=> Il existe une corrélation négative entre ces deux variables  Coefficient estimé : -0.5835339  (idem pour nombre de PAO max et min)
---	---

NB : les tests réalisés pour le milieu ouvert donne des résultats similaires (non montré)  
Comme pour le jeu de 39 espèces d'autres relations ont été testées mais n'ont pas donné de résultat significatif

## Résumé et Abstract

### Résumé :

Dans les communautés naturelles les modalités des traits des individus dessinent des patterns, qui peuvent être en partie le fruit de processus aléatoires mais aussi et probablement majoritairement construits suite à une sélection du milieu, selon la théorie des règles d'assemblage. Existe-il, chez les collemboles, des liens entre les caractéristiques du milieu et les patterns observés ? L'existence de ces liens a ici été étudiée à partir de la base de données COLTRAIT, regroupant des informations sur les différents traits morphologiques d'espèces de collemboles et sur les milieux où elles sont trouvées en Europe. Les analyses sur le jeu de données entier et sur des sous jeux de données ont montré l'existence d'association de modalités de traits et également l'existence d'une relation entre la présence/absence de l'organe post-antennaire (PAO), impliqué dans perception de l'environnement, avec la fermeture du milieu. Ces résultats ouvrent des perspectives sur le décryptage des règles d'assemblage propres aux collemboles dans leur milieu.

### Abstract :

Natural communities show pattern in the individual traits dispersion. These patterns may partly result from random processes or be the product of environmental selection, according to the assembly rules theory. Do links between environment characteristics and observed patterns exist in collembolan communities? The answer of this question was studied here, based on a database named COLTRAIT, which associates informations on morphologic traits from several species and environmental characteristics of where they are found in

Europe. Analyses on full data and sub-data samples show association between traits modalities, and a relationship between the presence/absence of the post-antennal organ (PAO) and environmental vegetation pattern. These results lead to Collembola assembly rules determination perspective.